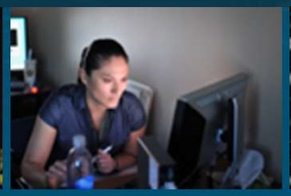
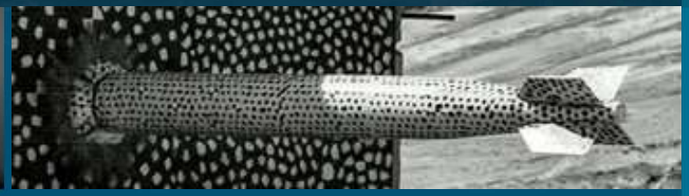
The 38th IEEE International Conference on Computer Design (ICCD 2020)
Special Session on Revisiting Adiabatic Circuits in the Era of Energy-Efficiency and Security

Special Session: Exploring the Ultimate Limits of Adiabatic Circuits

Sandia National Laboratories





Sunday, October 18th, 2020

Michael P. Frank, Center for Computing Research

with Robert W. Brocato, Thomas M. Conte, Alexander H. Hsia, Anirudh Jain, Nancy A. Missert, Karpur Shukla, and Brian D. Tierney

Approved for public release, SAND2020-10946 C

U.S. DEPARTMENT OF ENERGY | NNSA National Nuclear Security Administration

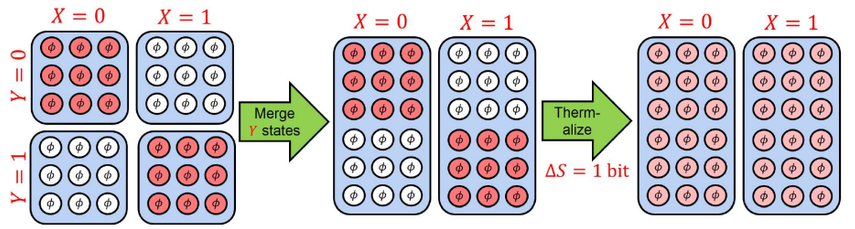
Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

Outline of Talk

Exploring the Ultimate Limits of Adiabatic Circuits

- I. Motivation & Brief History
- II. Technical Approach
  - Development of Fully Adiabatic CMOS Circuits
  - Fundamental Physics of Reversible Computing
  - Assessment of Architectural Implications
- III. Progress To Date and Early Results
  - Fully Adiabatic Logic Families: 2LAL and S2LAL
  - Simulation Studies
  - Test Chip Layouts
- IV. Next Steps
  - Important Upcoming Steps
  - Key Strategies for Leakage Reduction
- V. Conclusion

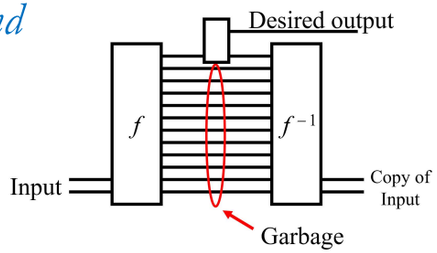
Motivation & Brief History



arXiv: 1901.10327

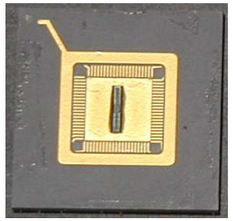
Landauer's Principle:

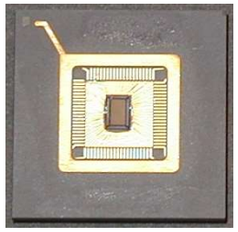
- Elementary statistical physics and information theory together imply that there is a *fundamental upper bound* on energy efficiency for the conventional (*non-reversible*) computing paradigm.
  - *Oblivious* erasure of known/correlated information implies dissipation of $E_{\text{diss}} \geq k_B T \ln 2$ energy to the environment for each bit's worth of known information that is lost.
    - k_B is Boltzmann's constant $\simeq 1.38 \times 10^{-23}$ J/K = the natural logarithmic unit of entropy.
    - T is ***the temperature of the thermal environment into which the waste heat ends up getting ejected.***
      - ∴ Simply lowering T *locally* cannot help *directly* to lower *system-level* E_{diss} if the *external* environment temperature is fixed.

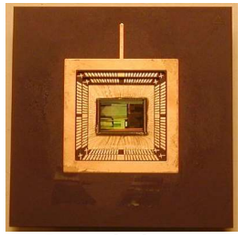


Reversible Computing (RC) provides a (theoretical, and maybe also practical!) solution:

- RC means computing *without* oblivious erasure of known or correlated information.
  - In principle, energy dissipation per useful operation can be made *arbitrarily small* (can approach zero as technology improves).
    - ∴ Energy *efficiency* (operations per Joule) can theoretically approach *infinity*!
  - This would have obvious implications for avoiding differential power analysis (DPA) and related side-channel attacks.


Tick


FlatTop


Pendulum

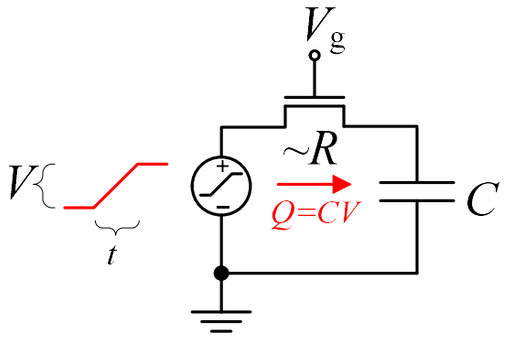
Brief history of the field:

- RC was first shown *theoretically* coherent by Bennett, 1973 (doi:10.1147/rd.176.0525).
- First *engineering* implementation proposed by Likharev, 1977 (doi:10.1109/TMAG.1977.1059351).
- First fully-adiabatic sequential CMOS logic style: Younis & Knight, 1993 (Proc. Int'l Symp. Res. Int. Sys.).
- First fabricated reversible processor chips! Frank, Knight, Love, Margolus, Rixner, Vieri (1996-1999).

The time is ripe for a resurgence!

- I believe there is an opportunity right now to demonstrate some real breakthroughs.

Adiabatic Charging via MOSFETs



A simple voltage ramp can *approximate* an ideal constant-current source.

- Note that the load gets charged up *conditionally*, if the MOSFET is turned on (gate voltage $V_g \gtrsim V + V_t$) during ramp.
  - V_t is the transistor's threshold, typically < ½ volt

Can discharge the load later using a similar ramp.

- Either through the same path, or a different path.

$t \gg RC \Rightarrow E_{diss} \to CV^2 \frac{RC}{t}$

$t \ll RC \Rightarrow E_{diss} \to \frac{1}{2} CV^2$

Exact formula for linear ramps:

$$E_{diss} = s\left[1 + s\left(e^{-1/s} - 1\right)\right]CV^2$$

given *speed fraction* $s = RC/t$.

The (ideal) operation of this circuit approaches *physical reversibility* ($E_{diss} \to 0$) in the limit $t \to \infty$, but *only* if a certain *precondition* on the initial state is met (namely, $V_g \gtrsim V_{max} + V_t$)

- How does the possible physical reversibility of this circuit relate to its *computational* function, and to some *appropriate* concept of logical reversibility?
  - Traditional (Landauer/Fredkin/Toffoli) reversible computing theory does **not** adequately address this question, so, we need a more powerful theory!
  - The theory of ***Generalized Reversible Computing*** (GRC; mentioned briefly yesterday) meets this need.

See arxiv:1806.10183 for the full GRC model.

Trapezoidal Resonators via Fourier Decomposition

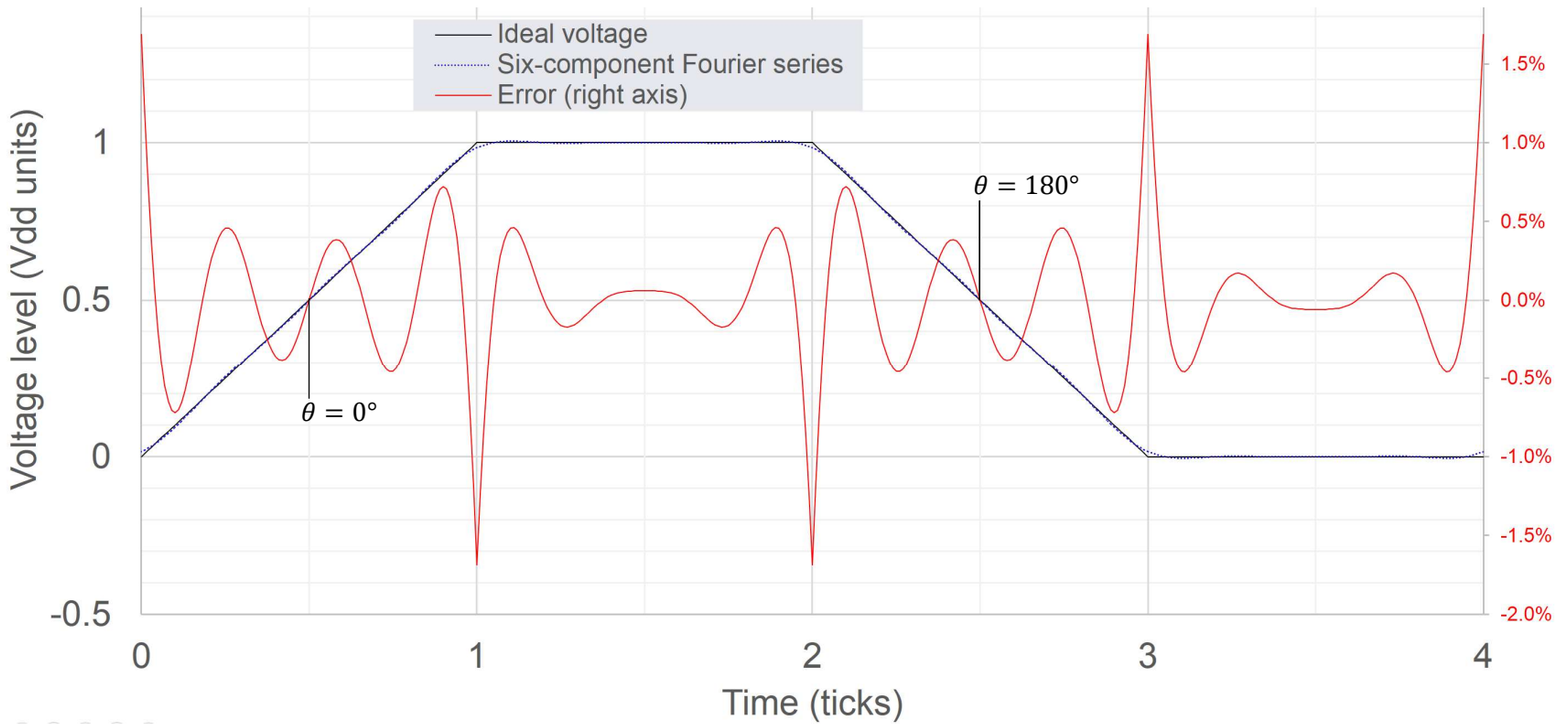
E.g., consider the ideal trapezoidal waveform shown below

Note, relative to mid-level crossing, waveform is an *odd* function
- ∴ Spectrum includes only odd harmonics $f, 3f, 5f, \ldots$

Six-component Fourier series expansion is shown below
- Maximum offset with $11f$ frequency cutoff is $< 1.7\%$ of V_{dd}

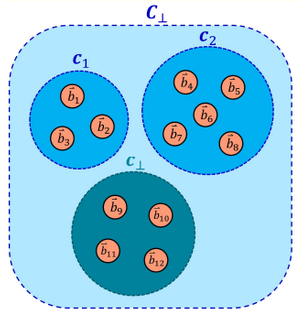
S. G. Younis and T. F. Knight, "Non-dissipative rail drivers for adiabatic circuits," in Proc. Sixteenth Conf. on Advanced Research in VLSI, IEEE, Mar. 1995, pp. 404-414. doi:10.1109/ARVLSI.1995.515635

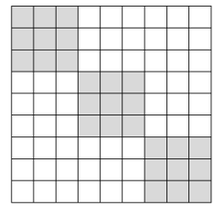
$$v_{f6}(t) = V_{\mathrm{DD}}\left[\frac{1}{2} + \frac{4\sqrt{2}}{\pi^2}\left(\begin{array}{l}\sin(\omega t) + \\ \frac{\sin(3\omega t)}{3^2} - \\ \frac{\sin(5\omega t)}{5^2} - \\ \frac{\sin(7\omega t)}{7^2} + \\ \frac{\sin(9\omega t)}{9^2} + \\ \frac{\sin(11\omega t)}{11^2}\end{array}\right)\right]$$

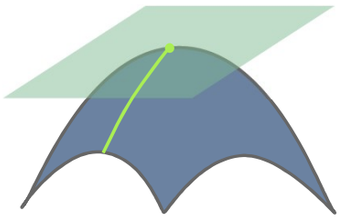


University Collaborations

$$U_S^t(\mathfrak{S}, \mathcal{B}) \Vdash C_S^t(O_S^t, \rho_S)$$





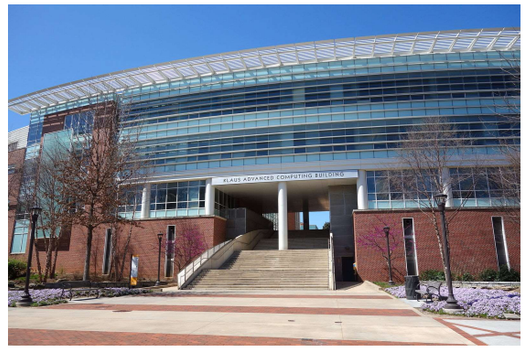


Karpur Shukla with Jimmy Xu (Brown U.)

- Fundamental physics of reversible computing
  - Limits on dissipation as a function of delay, etc., from nonequilibrium quantum thermodynamics?
  - Harnessing exotic quantum phenomena to improve efficiency of reversible devices

Tom Conte and Anirudh Jain (Georgia Tech)

- Architectural implications of improved reversible technologies
  - Up to the level of full processor designs
  - Looking at high-performance computing applications
- Optimization of energy/cost tradeoffs in partially-reversible designs



Fully Adiabatic Logic Families: 2LAL & S2LAL

To approach ideal reversible computing in CMOS…

We must aggressively eliminate *all* sources of non-adiabatic dissipation, including:

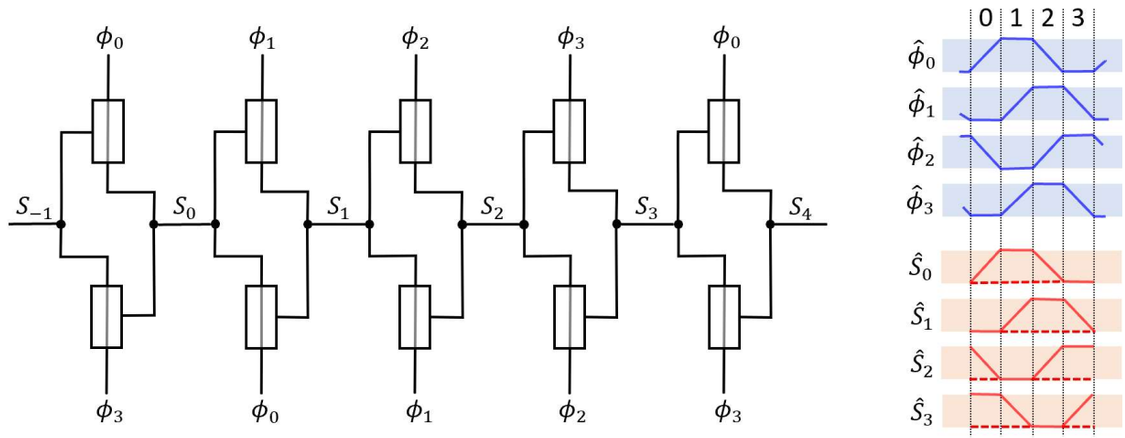
- Diodes in charging path, "sparking," "squelching,"
  - Eliminated by "**truly, fully adiabatic**" design. (*E.g.*, CRL, 2LAL).
    - Suffices to get to a few aJ (10s of eV) in 180 nm *before* voltage optimization.
- Voltage level mismatches that *dynamically* arise on floating nodes before reconnection.
  - Eliminated by static, "**perfectly adiabatic**" design. (*E.g.*, S2LAL).

We must also aggressively minimize standby power dissipation from leakage, including:

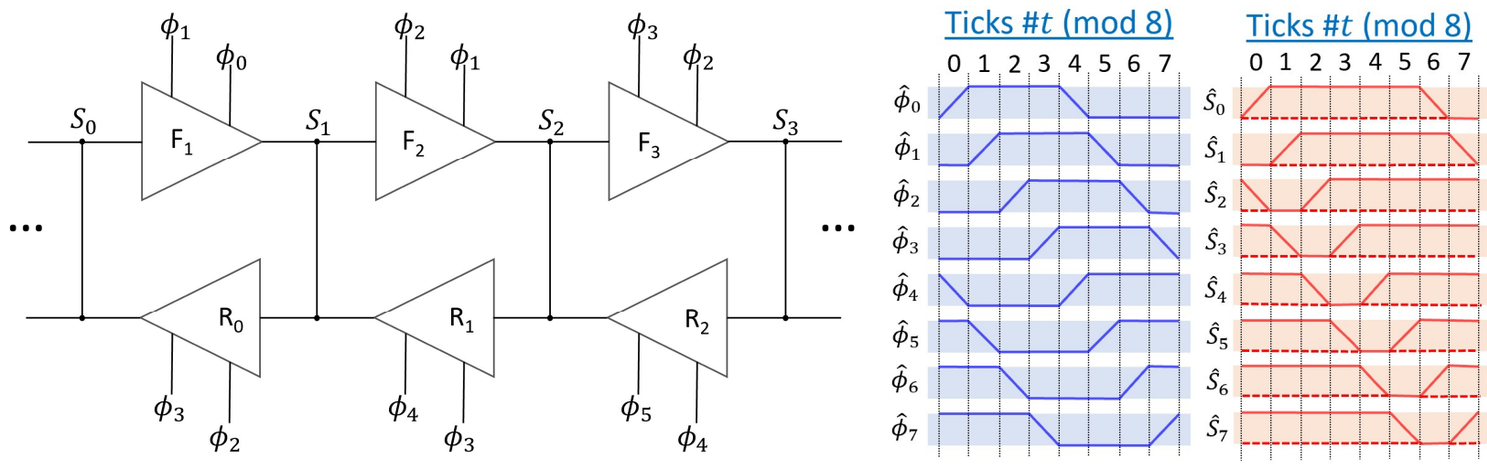
- Subthreshold channel currents
  - Low-*T* operation helps with this
- Gate oxide tunneling
  - *E.g.*, use thicker gate oxides

Note: (Conditional) logical reversibility *follows from* perfect adiabaticity.

Shift Register Structure and Timing in 2LAL



Shift Register Structure and Timing in S2LAL



(arxiv:2009.00448)

Simulation Studies

Processes studied to date:

- Sandia MESA (350 nm and 180 nm)
- GlobalFoundries 180 nm RF-SOI

We have been utilizing a number of different tools in our simulation studies:

- Compact model (SPICE) based circuit simulators (using BSIM3, BSIM3SOI, BSIM4 based models):
  - Cadence/Spectre
  - Keysight Pathwave ADS (Advanced Design System)
  - ngspice
- Technology CAD / finite-element device simulators:
  - Silvaco TCAD
  - (Future) Synopsys Sentaurus
  - (Possible Future) Sandia Charon

Snapshot of some representative results from simulation studies to date:

- Energy dissipation per cycle per FET in 2LAL shift register at 50% activity factor at $f = 1$ MHz, $C_L = 10$ fF:
  - Spectre simulation of MESA 350 nm, $W = 800$ nm: **37 aJ ≈ 230 eV.**
  - Spectre simulation of MESA 180 nm, $W = 480$ nm: **6.9 aJ ≈ 43 eV.**
    - Comparable to data at the same operating point for TSMC18 from a 2004 study of 2LAL at UF.

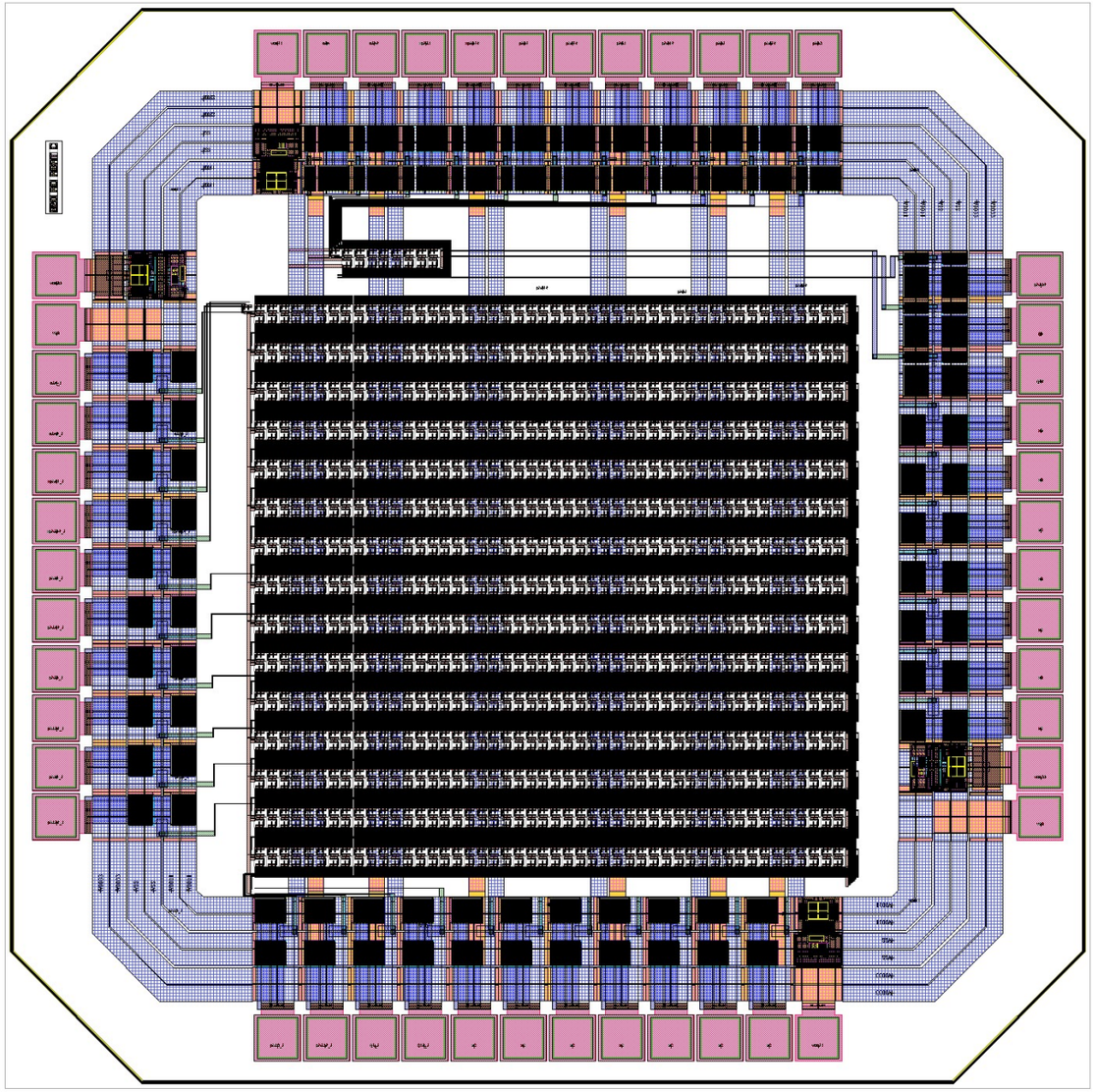
- We expect these results can be substantially further improved by optimizing over the f, V_{DD}, V_{SB} parameter space.

Test Chip Layouts

Sandia MESA 180 nm CMOS process
- Shuttle run, taped out August 2020
- Die size: 2×2 mm

2LAL shift registers
- 8 stage and 720-stage

Goals for test chip:
- Verify shift register functionality
- Empirically measure energy dissipation



Important Near-Term Next Steps

Further optimization of simulated results for 2LAL.

- Exploration of parameter space.

Simulations, layouts and test chip fabrication for the newly invented fully static version of 2LAL (S2LAL).

Further development of the trapezoidal resonant oscillator circuit.

- Including experimental verification that prototype implementations can drive the 2LAL test chip,
- And measurements of system power dissipation (resonator coupled to test chip).

Exploration of additional available processes as implementation candidates.

- Including possibly Skywater, MIT LL, MOSIS processes.

Possible step:

- Development of an *open source* cell library for fully/perfectly adiabatic CMOS.

Further Reducing Dissipation in the Long Term

Two key approaches to this:

- *Use older processes.* (Counterintuitive, but true!)
  - Note that *both* subthreshold leakage and gate leakage scale down *exponentially* as dimensions and voltages increase.
  - This overwhelms the merely *polynomial* impact of having larger C and V values.
  - ∴ Minimum energy dissipation of adiabatic circuits in older processes, when well optimized, should be lower than in newer ones. (At room temperature.)
- *Operate at cryogenic temperatures.*
  - Steeper subthreshold slope → much lower subthreshold leakage.
    - Can re-optimize device structures for cryo to bring gate leakage down to meet subthreshold leakage at lower total levels.
  - Resonators can be built from superconducting materials, or moved out to room-temperature environment.
  - Superconducting *interconnects* can help reduce the wiring component of the R term in the CV^2RC/t expression.
  - Such methods should allow newer processes operated at cryo to exceed the energy efficiency of older ones, and at faster speeds.
    - *Dissipation-delay product* will be lower in newer processes *at cryo* than in older processes.

Conclusion

The limits of energy efficiency achievable with adiabatic CMOS are very far from being reached!
- No fundamental limit to the efficiency of this approach is yet known.

Some necessary steps for continuing to make progress:
- Utilize *truly, fully adiabatic* reversible logic styles,
- Beyond that, utilize fully-static, *perfectly adiabatic* reversible logic styles.
- Develop high-Q trapezoidal resonant oscillators.
- Minimize leakage aggressively, *e.g.*, by re-optimizing the process for adiabatic operation at cryogenic temperatures.

In the long run, as the efficiency gains achievable through adiabatic operation continue to increase,
- There will be increasing demand for fabrication processes that stack multiple layers of active devices in 3D.
  - Because doing so will reduce interconnect-related overheads substantially.
- Thus, the cost per-layer (and per-device) for 3D active logic processes will also continue to decrease.
- Overall system cost-efficiency including *both* energy-related costs and manufacturing costs can continue improving.
  - Note, however, that this would not be possible at all, beyond some point, if we don't approach fully adiabatic and reversible operation!

In the very long term, exploration of the fundamental quantum physical limits of reversible computing can lead to new device-level breakthroughs.
- Leading to further dramatic improvements in efficiency.

Due to Landauer's Principle, the vastly more efficient computers of the far future will have to be reversible!